基于模仿学习的机场停机位再分配决策算法 *

邢志伟 [1]，张前前 [1]，罗　谦 [2†]，陈肇欣 [2]

(1. 中国民航大学 电子信息与自动化学院，天津 300300; 2. 中国民用航空局第二研究所 工程技术研究中心，成都 610041)

摘　要：针对机位再分配算法结果难以满足不同操作人员操作习惯的问题，提出一种符合实际业务人员操作习惯的机位再分配推荐算法。首先以航班特征属性和停机位的资源占用状态构建决策环境空间模型，将人工操作数据转换为多通道时空矩阵，再以卷积神经网络构建的生成对抗网络(generative adversarial network，GAN)拟合其序贯决策操作策略。仿真结果表明，可靠度在 90%以上的调整动作占比最高达到 84.4%。经过在 3 个数据集上的测试，模型对不同来源的操作数据具有较好的区分能力。对比不同扰动下的动态调整结果，算法能够得到航班-机位属性特征与原有人工操作属性特征接近的调整方案。

关键词：航空运输；停机位分配；模仿学习；马尔可夫决策过程；生成对抗网络

中图分类号：V351　　doi: 10.19734/j.issn.1001-3695.2022.02.0071

Decision-making algorithm for airport gate reassignment based on imitation learning

Xing Zhiwei[1], Zhang Qianqian[1], Luo Qian[2†], Chen Zhaoxin[2]

(1. *School of Electronic Information & Automation*, *Civil Aviation University of China*, *Tianjin* 300300, *China*; 2. *Engineering Technology Research Center*, *The Second Research Institute of Civil Aviation Administration of China*, *Chengdu* 610041, *China*)

Abstract: In order to solve the problem that the results of the gate reassignment algorithm can't meet habits of different operators, this paper proposed a method that accords with the actual operators' operating habits. Firstly, this paper established the spatial model of decision-making environment by using flights characteristics and occupancy of gate resources. The model transforms manual operating data into a multi-channel time-space matrix. Then, make use of CNN-based generative adversarial network to match the order decision-making operation strategy. The simulation results show that actions with reliability scores of more than 90% account for up to 84.4%. The model has a good ability to distinguish the operation data from 3 different operators. Compared with dynamic adjustment result under perturbance, this algorithm can obtain an adjustment scheme whose flight-gate attribute characteristics are closer to the original manual operation.

Key words: air transportation; airport gate assignment; imitation learning; Markov decision process; generative adversarial network

0　引言

停机位分配算法是辅助机场调度人员完成航班停机位指派的重要技术手段，该问题作为机场运行控制领域重要的研究方向持续受到关注[1]。

现有的停机位预分配算法研究大多采用动态规划的思路进行算法建模[2]，停机位再分配算法通常以最小化与原计划之间的变化为优化目标[3~9]。Deng[3]通过遗传-蚁群混合算法(genetic algorithm and ant colony optimization, GA-ACO)最小化乘客、机场和航司的经济损失。Zhang[4]构建机位再分配为网络流模型并通过变滚动层位算法求解。姜雨[8]等在对航班延误后实时调度问题的研究中考虑到延误的等级划分，在不同的延误严重程度下，针对不同的优化目标分别优化，使用非支配排序遗传算法求解，得到差异化多目标机位分配模型。Maharjan[9]等提出了一种基于传统确定性模型的机位动态分配模型，建立了最小化旅客的传递距离的二次整数规划模型。

根据上述的研究，既有的机位分配算法会综合考虑航班靠桥率、鲁棒性等若干复杂多样的优化目标，算法模型构建过程中往往通过人为权重参数分配或优先级设定方式平衡各优化目标之间的关系以进行动态规划求解。然而实际运用中对于大量优化目标或约束条件，业务人员很难主观精确量化符合机场实际运行目标的算法参数配置，导致算法计算得到分配结果后仍然需要人工进行核查及再分配调整。

模仿学习(Imitation Learning，IL)，也称为示例学习，通过相对简单的方法模仿某种行为策略作出反映。它从数据集中提取有用的知识，在类似于演示的环境中重现其行为策略以实现与之类似的决策方式。IL 的存在有助于自主控制系统的研究和人工智能策略模型的设计，因为此类算法在现实场景中表现出良好的前景和策略训练的效率。在连续控制领域，模仿学习可应用于自主车辆操纵，以在动态环境中再现适当的驾驶行为[10,11]、机器人控制[12~14]等。在离散控制领域，模仿学习在导航任务[15~17]等领域作出了贡献。一般而言，算法采用的演示数据是从人类专家或人工代理与环境

收稿日期：2022-02-13；**修回日期：**2022-04-02　　**基金项目：**国家重点研发计划资助项目(2018YFB1601200)；四川省青年科技创新研究团队专项计划资助项目(2019JDTD0001)；四川省科技计划资助项目(2021003)；成都市重点研发支撑计划资助项目(2019-YF08-00265-GX)

作者简介：邢志伟(1970-)，男，辽宁沈阳人，教授，硕导，博士，主要研究方向为民航装备与系统、民航智能规划与调度、机场运行控制与信息；张前前(1997-)，男，湖南衡阳人，硕士研究生，主要研究方向为机场运行控制与信息；罗谦(1975-)，男(通信作者)，四川宜宾人，研究员，硕导，博士，主要研究方向为机场运营管理、数据挖掘(caacsri_luoqian@163.com)；陈肇欣(1988-)，男，江苏南京人，中级工程师，博士，主要研究方向为机场运营管理、数据挖掘.

交互中收集获得。

IL 按惯例分为行为克隆(Behavioural Cloning，BC)、逆向强化学习(Inverse Reinforcement Learning ,IRL)和对抗结构模仿学习(Adversarial Structured Imitation Learning ,ASIL)。BC 和 IRL 方法使用不同的方法来产生专家行为。一般而言，通过 IRL 构建得到了奖励(成本)函数后，需要进一步采用该奖励函数代入 RL 中得到被学习的专家策略，这种间接的方式在较小的空间中求解是可行的，而当环境空间进一步扩大时，将会面临求解成本高昂的问题。而行为克隆算法直接通过专家轨迹数据得到相应的策略，但该算法会在专家轨迹未出现的状态上产生偏移问题，所以需要大量的训练数据，不适合数据获取成本较高的应用场景。ASIL 最初产生于由 Ho 和 Ermon 提出的 GAIL 算法[18]，尽管 ASIL 与 IRL 有着密切的联系，但大多数对抗性结构 IL 都不能恢复奖励函数。BC 通常利用在环境中没有交互作用的专家之间的交互作用，故 BC 的计算成本是最低的；IRL 方法通常在其算法内环中与环境有着较多的交互作用，对系统动态的评估使 IRL 计算量大；ASIL 方法在迭代更新策略参数和鉴别器参数时，还涉及与环境的频繁交互。

人工再分配调整本质是一种决策策略，与动态规划算法模型对某些航班的决策存在偏差，这种偏差产生于模型参数的主观设置。为了消除这种偏差并更准确地刻画人工的决策过程，将人工再分配调整建模为序贯决策问题，提出了一种基于特征嵌入的生成对抗模仿学习方法(feature embedding-generative adversarial imitation learning，FE-GAIL)的停机位再分配决策模型，通过直接学习人工操作数据，解决静态参数下的人工主观参数配置不准确的问题。

1　航班停机位再分配模型设计

1.1　问题描述

机场通常提前一天获取次日航班时刻表，经过机位分配算法计算得到的预分配计划以甘特图(图 1)的形式呈现。在当天实际运行时，航班时刻变化可能导致机位使用冲突，算法会在该计划上进行再分配以消解冲突。

航班的机位调整在同一场景下往往存在多种可行解。如图 1 所示，当 103 机位上的航班 5 发生延误时，将与航班 6 产生冲突，需要将航班 6 调整至其他机位。图中①、②和③分别代表了 3 种可行的调整方案，传统多目标优化的算法得到最优解①。

经过对机场操作人员的调研，不同的操作人员对传统算法中配置参数方案的认知是不一致的，这导致实际业务中难以给定一个具有广泛适用性的参数配置，算法输出的结果中不同的操作人员可能选择不同的推荐方案，甚至其选择并非算法输出的最优解。

例如，设 103 机位为远机位，101 机位为廊桥，操作人员 1 倾向于调整更少航班数量而选择了方案②，而操作人员 2 倾向于更高廊桥靠桥率选择了可行解③。

这种算法计算结果和员工实际行为上的偏差来自于个人主观经验与参数主观设置的偏差，传统的多目标优化算法难以通过参数优化设置方式消除这种偏差。针对上述问题，通过在多目标优化过程中引入机器学习算法，实现主观设置参数的偏差纠正。再分配过程中，一个航班机位的人工调整可能需要多个航班机位的连锁调整以得到可行解，因此停机位再分配可以视为一个序贯决策过程，针对序贯决策的策略学习问题通常采用模仿学习算法[19]。其中生成对抗模仿学习(Generative Adversarial Imitation Learning，GAIL)算法主要通过示例决策轨迹数据学习其中的决策策略，得到近似于示例的决策策略，在自动驾驶等序贯决策场景具有较好的状态泛化效果[20]。综上，基于 GAIL 算法的思想，将航班与机位的状态特征信息进行特征值嵌入操作(feature embedding，FE)，通过学习人工操作轨迹数据来拟合人工停机位再分配的决策过程，以消除主观经验与参数主观设置的偏差。

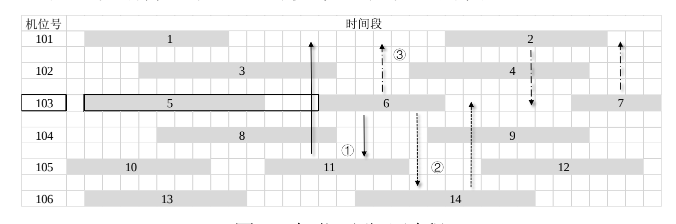


图 1　机位再分配过程

Fig. 1　Process of gate reassignment

图 2 给出了基于 FE-GAIL 的停机位分配策略学习算法流程。首先通过航班计划表和停机位属性构建决策环境空间，之后设计基于该决策环境空间下的神经网络用以决策输出。

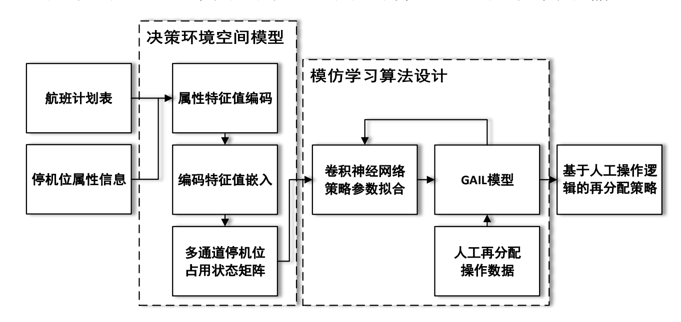


图 2　基于 FE-GAIL 的停机位分配策略学习算法

Fig. 2　Algorithm of gate reassignment based on FE-GAIL

1.2　决策环境空间模型

为更好地描述决策环境空间模型，首先定义相关概念：

定义 1　待分配航班 f_o：指当前需要被调整机位的航班。

定义 2　动态滑动窗口 T_w(min)：指再分配过程中划定需要考虑的时间范围的窗口。停机位的再分配问题中可调整对象一般为落地时间晚于待分配航班 f_o 的其他航班，起始点是待分配航班 f_o 的实际落地时间。

GAIL 是一种基于马尔可夫决策过程(markov decision process, MDP)的算法，根据各个时间点观察环境得到的状态，选择需要执行的动作，并根据环境返回值循环往复得到执行动作序列。机位再分配调整过程中，交互环境模型以及状态转移概率处于未知状态，随参数的迭代而迭代更新，故将 MDP 描述为一个四元组 $\langle S,A,R,\gamma\rangle$ 。其中 S 为状态空间即 $S=\{s_1,s_2,\ldots,s_i,\ldots\}$ 。表示某一时刻 t 下的机位使用计划的特征空间描述 s_i 。动作空间 A 是可以执行的动作 a 的集合，即 $A=\{a_1,a_2,\ldots,a_i,\ldots,a_N\}$ ，其中 a_i 表示操作人员将待分配航班 f_o 的调整至机位 i 。 R 为奖励函数，而 γ 表示学习率。

1.2.1 多通道机位状态特征矩阵

为定义机位状态空间 s_t ，首先需要明确机位分配的决策因素。影响人工机位再分配决策的因素主要包括机位特征(机型尺寸属性、业务属性、航空公司属性)、航班特征(机型尺寸属性、业务属性、航空公司属性)、待分配航班属性(同航班特征)及机位前后时段占用状态等，因此本文建立了如图 3 所示的机位状态环境特征模型，将动态滑动窗口内的机位甘特图映射为状态特征矩阵。

图 3 左侧为停机位占用状态的甘特图，其中每行对应于一个机位，每列对应于一个时间块。右侧为针对停机位再分配中特征信息提出的多通道机位状态特征矩阵，其中每个通道表示一种航班属性或停机位属性所对应的嵌入编码，具体的编码嵌入方式在 1.2.3 节中进一步展开叙述。

假设停机位总数量为 N ，为了减小状态维度，以 10 分

钟为一个时间块，动态滑动窗口中需要表征的时间块数量为 $T_b = T_w/10$，故占用矩阵每个通道的矩阵维度大小为 $T_b \times N$。

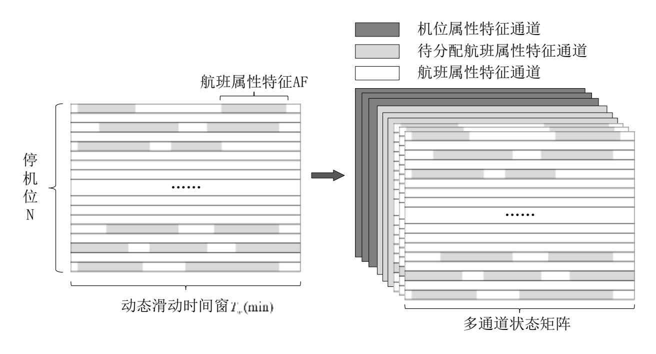


图 3　多通道机位占用状态矩阵

Fig. 3　Multi-Channels occupation state matrix

1.2.2 属性特征

航班和停机位的特征是决定航班和停机位匹配关系的重要因素，表 1 给出了部分典型的属性特征说明。

表 1　属性特征表

Tab. 1　Attribute features list

属性名称	属性值	属性说明
机型尺寸属性	A/B/C/D/E/F	航班的机型尺寸不得超过停机位尺寸
业务属性	国际/国内/地区	国际航班应停靠在国际机位，国内亦同，部分远机位无限制
航空公司属性	CA、CZ 等	航班应停靠对应航司归属停机位

以航班-机位尺寸匹配属性为例，传统的停机位分配算法中需要人工设置匹配关系的优先级权重表(如表 2 所示)，以尽可能减少小飞机占用大机位的情况。为了避免大量属性权重值的人为给定，通过将航班和停机位的属性以特征编码的形式建立状态特征矩阵，其权重匹配关系不显性地体现在算法模型中，而是通过神经网络的参数表示，再通过 GAIL 的方式进行参数学习。

表 2　航班-机位尺寸匹配权重

Tab. 2　Flight-gate size matching weight

飞机\机位	A 类	B 类	C 类	D 类	E 类	F 类
A 类	100	80	60	40	20	10
B 类	/	100	60	40	20	10
C 类	/	/	100	70	50	30
D 类	/	/	/	100	80	60
E 类	/	/	/	/	100	80
F 类	/	/	/	/	/	100

根据表 1 的属性特征表，设航班集合为 F，则第 k 个航班的属性以向量的形式表示为

$$A_{fk}=[a_{f1k},a_{f2k},a_{f3k}] \tag{1}$$

其中 a_{f1k} 表示航班 k 的机型尺寸属性，a_{f2k} 表示业务属性，a_{f3k} 表示航空公司属性。

第 i 个停机位的属性向量表示为

$$A_{gi}=[a_{g1i},a_{g2i},a_{g3i}] \tag{2}$$

其中 a_{g1i} 表示停机位 i 机位尺寸属性，a_{g2i} 表示业务属性，a_{g3i} 表示航空公司属性。

1.2.3 属性特征的编码嵌入

由于航班属性特征与机位的占用情况相关，故首先定义机位的占用状态特征矩阵。设图 3 中停机位占用状态特征矩阵表示如下：

$$\boldsymbol{O}=\begin{bmatrix} o_{00} & o_{01} & \cdots & o_{0T} \\ o_{10} & o_{11} & \cdots & o_{1T} \\ \vdots & \vdots & \ddots & \vdots \\ o_{N0} & o_{N1} & \cdots & o_{NT} \end{bmatrix} \tag{3}$$

其中 $o_{ij}\in\{0,1\}$ 表示第 i 个机位在第 j 个时间块上的决策变量，若航班 k 停靠于机位 i 则取 1，反之取 0。

根据 1.2.2 节中定义的属性特征，需要将属性值对应成编码的形式嵌入到机位占用状态矩阵的元素中，即特征嵌入(Feature Embedding，FE)。停机位和航班的相对关系与电子游戏中的地图与可操作单位的相对关系具有一定的相似性，Oriol Vinyals[21]等将 StarCraft II 中的小地图信息进行基本编码，而可操作单位用特征向量进行表征并放置于地图中，从而得到信息编码后的完整状态。

根据上述特征嵌入(FE)方式，首先对属性值进行编码。将每种属性值单独编码成为一个通道，以机型尺寸属性为例，机型尺寸属性特征值可取 $\{A,B,C,D,E,F\}$，按照等间距的方式取值，若机位在某一时间块被机型尺寸属性为 D 的飞机占用(属性排序为 4)，则该飞机对应占用的时间块的属性值取 $4/6=0.67$。

再将编码后的属性值嵌入到状态矩阵中。航班第 v 种属性对应的通道中第 k 个航班对应的元素可以表示如下：

$$x_{ijk}=a_{fvk}\cdot o_{ij} \quad (0\le i\le N,0\le j\le T,k\in F) \tag{4}$$

根据上式的表述，当某个时间块上没有航班占用，则对应值为 0，反之则为属性值。

对于停机位属性通道，则无须表征占用状态特征，第 u 种属性对应的通道中，第 i 个停机位所对应的矩阵元素表示为如下形式：

$$x_{iju}=a_{gui} \quad (0\le i\le N,0\le j\le T) \tag{5}$$

完成所有航班的属性值嵌入后，最终得到该属性对应的矩阵通道如图 4 所示，其他属性对应的矩阵通道的构建方式同理。

	t_1	t_2	t_3	···	t_{T-1}	t_T
g_1	0.67	0.67	0.67	···	0	0
g_2	0	1	1	···	1	0
g_3	0	0.33	0.33	···	0	0
⋮	⋮	⋮	⋮		⋮	⋮
g_{N-1}	0.17	0.17	0	···	0	0
g_N	0	0	0	···	1	1

图 4　状态矩阵中机型尺寸属性通道

Fig. 4　Flight size attribute channel in the state matrix

2　模仿学习算法设计

模仿学习是一种基于序贯决策过程的人工数据学习决策策略算法(Feature Embedding-Generative Adversarial Imitation Learning，FE-GAIL)，本章将说明 FE-GAIL 算法在停机位再分配决策场景的算法设计。

按照上述业务场景的建模，拟采用生成对抗模仿学习(GAIL)求解机位再分配策略学习问题，为了便于描述数据的特性，在说明算法流程之前先定义相关概念——占用度量 ρ (occupancy measure)。

定义 3　占用度量 ρ：使用策略 π 时在环境中得到的状态-动作对 (s_i,a) 的分布，且占用度量 ρ 和策略 π 之间存在唯一的对应关系[18]：

$$\rho_\pi(s_i,a)=\pi(a|s_i)\sum_{t=0}^{\infty}P(s_t=s_i|\pi) \tag{6}$$

其中，$\pi(a|s_i)$ 和 $P(s_t=s_i|\pi)$ 分别表示动作和状态的边缘分布。根据上述定义可知，当策略与人工停机位再分配策略的占用度量越接近时，两个策略之间越相似。将 GAN 结合到正则优化器 φ 中，将优化器定义为

$$
\begin{aligned}
&\varphi(\rho_{\pi},\rho_{E})= \\
&\text{maxmize}\sum_{s,a}\rho_{\pi}(s,a)\ln(D(s,a))+ \\
&\rho_{\pi_E}(s,a)\ln(1-D(s,a)) \\
&=\text{maximize } E_{\pi}[\ln(D(s_i,a))]+E_{\pi_E}[1-D(s_i,a)]
\end{aligned}
\tag{7}
$$

其中 D 为生成对抗网络中的鉴别器。生成器 π 生成用于混淆鉴别器 D 的样本，而 D 则用于区分和人工策略样本。

将策略 π 的占用度量类比为生成器 π 生成数据的概率分布，人工策略类推。此概率分布可以刻画为选择状态下的动作概率分布向量 $\boldsymbol{v}_{\pi}=[p_{a1},p_{a2},...,p_{aN}]$，$p_{ai}\in[0,1]$，其中的元素为在给定状态下各动作被选择概率。当鉴别器 D 无法区分两者时，认为 π 具备产生与人工样本相近似样本的能力。

算法流程图如图 5 所示。

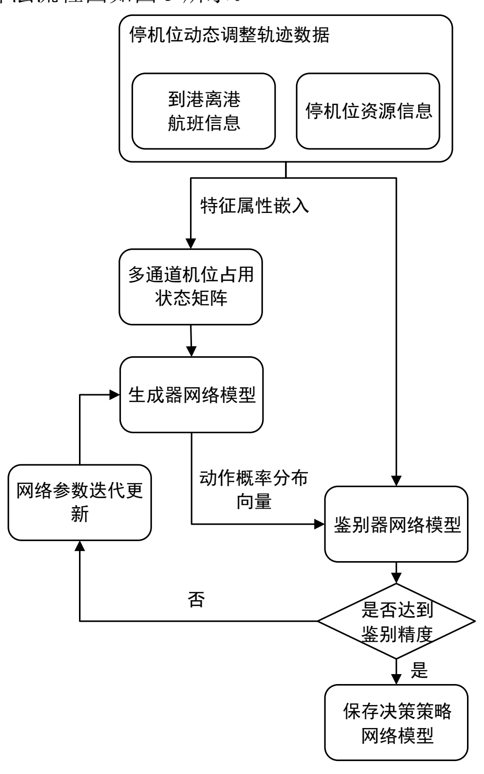


图 5 FE-GAIL 算法流程图

Fig. 5 FE-GAIL algorithm flow chart

基于 FE-GAIL 的停机位再分配决策算法的基本训练流程如下：

Setp1 初始化生成器 π 和鉴别器 D 的网络参数，设置训练批量大小；

Step2 将人工停机位再分配操作数据以状态-动作对 (s_i,a) 进行特征嵌入(FE)作为生成器 π 的输入，得到选择机位的动作概率分布向量 v_{π}；

Step3 按照动作概率分布向量 v_{π} 选择机位，得到生成器的输出 (s_i,a_{π})；

Step4 将人工停机位再分配操作数据和生成器产生的样本同时输入鉴别器 D 中，得到的鉴别结果根据式(7)，利用梯度更新鉴别器 D 和生成器 π 的网络参数 θ。

通过上述算法，最终得到能够满足需求的生成器 π 和鉴别器 D，其中生成器即为策略模型，而鉴别器将在 3.3 节中用于测试模型区分度。

3 实验分析

本章将通过数据实验的方式证明根据停机位再分配特性提出的算法可以学习到人工再分配决策策略。下文将首先介绍实验所用的数据，将通过三类实验对比 IRL 和 BC 算法，验证算法的性能。

3.1 实验数据

根据 1.1 节可知，机场不同的操作人员具有各自不同的算法调参策略，即对统一参数有不同的参数配置方式。在此假设同一员工在进行行为决策时始终遵循相同的行为逻辑，利用基于 A*的启发式搜索算法模拟人工决策过程，以扩充原生较少的人工操作数据，生成足够训练算法模型的行为决策数据。

A*算法通过判断当前位置与起点之间的距离以及与终点之间的预估代价来求解代价最小的可行解，适用于多步决策下的最优决策路径搜索问题。其代价预估函数计算式如下：

$$F(s)=G(s)+H(s) \tag{8}$$

其中 $G(s)$ 为当前路径的实际代价，$H(s)$ 为行为预估代价。对于第 i 个停机位和第 k 个航班，$G(s)$ 可以表示如下：

$$G(s)=\phi(A_{fk},A_{gi},A_{pki}) \tag{9}$$

其中，A_{fk} 和 A_{gi} 分别为由式(1)和(2)得到的航班和机位属性，A_{pki} 为根据表 2 的形式得到的航班-机位匹配属性。

按照上述权重计算方式，将待分航班可选机位作为可扩展的相邻节点，通过路径搜索的方式得到如图 1 所示的动态调整动作序列。

实验采用国内某机场中的 40 个停机位，分为 4 个指廊，共计 45700 余条原计划时间表的真实数据，如表 3 所示。

表 3 航班-停机位原始计划表

Tab. 3 Original schedule of flight-gates

航空公司	航班属性	机型	机位	起飞时间	落地时间
1	国际	333	103	5/19 02:34	5/18 23:30
2	国内	773	128	5/19 10:47	5/18 23:33
3	国际	321	101	5/19 1:24	5/18 23:34
4	国际	738	102	5/19 7:18	5/18 23:52
······	······	······	······	······	······

在此基础上对航班落地和起飞时间加入随机扰动，用于模拟机场发生延误时的实际变动情况，再利用 A*算法模拟人工分配的决策行为，以得到人工分配操作行为轨迹数据。根据历史延误信息，将航班的延误分布情况用泊松分布描述：

$$P(x=k)=\frac{\lambda^k}{k!}e^{-\lambda} \tag{10}$$

其中 k 为延误时间块的数量。定义航班 f_k 与停机位 g_i 对应的权重参数集 W 如下所示。

$$W=[A_{fk},A_{gi},A_{pki}] \tag{11}$$

根据策略不变假设，对于员工 i 而言其行为参数集 W 始终保持不变。为了测试模型对于不同员工行为逻辑的学习能力和区分度，本文设置了三组参数集 $W_i, i\in\{1,2,3\}$ 以模拟三种不同的分配操作行为逻辑，利用上述的 A*算法生成三组数据集 $D_i, i\in\{1,2,3\}$，每组数据集 D_i 包括训练数据 10000 条，测试数据 2000 条。

3.2 模型可靠度结果分析

为了量化训练后的模型分配结果与人工操作行为的一致性水平，根据 **2.1** 节中的动作概率分布向量 $\boldsymbol{v}_{\pi}$，本文定义归一化的可靠度评分 R 如下：

$$r_i=\begin{cases}1, \text{ if } p_{ai}\le p_{a_E}, \ p_{a_E},p_{ai}\in v_{\pi}, \\ i\in[1,N] \\ 0, \ otherwise\end{cases} \tag{12}$$

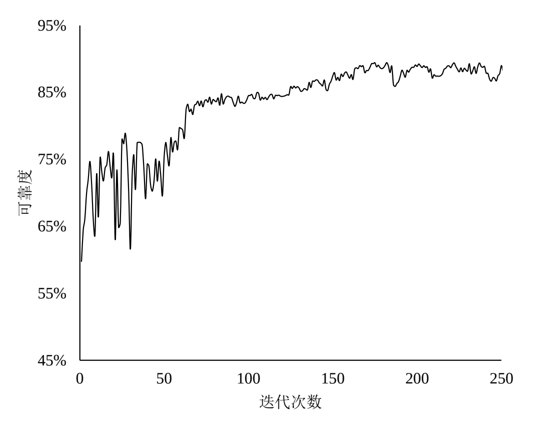
$$R=\frac{1}{MN}\sum_{i=0}^{M}\sum_{i=0}^{N}r_i \tag{13}$$

其中 M 为每次测试采样的样本数量，N 为停机位数量，a_E 为人工操作动作。

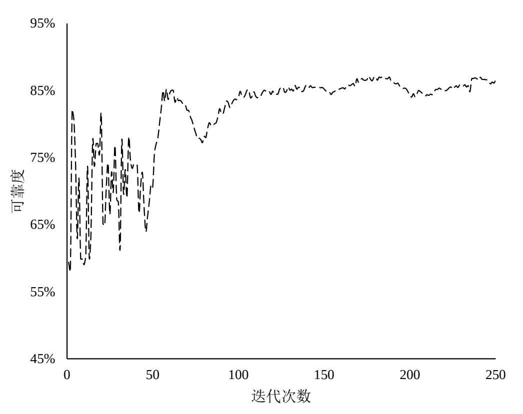
取动态滑动窗口时间长度 T_w 为 300 分钟，对应的时间块数量为 30，状态空间维度为 40*30*9。训练过程中每训练 100 次作为一次迭代，并进行一次测试，每次测试采样的人

工轨迹样本数量为 1000，设置学习率 $\gamma=0.001$，测试 10 轮得到平均可靠度评分的变化趋势图如图 6 所示，其中三条曲线分别表示来自三名不同操作人员给定的权重参数所对应的数据集进行 FE-GAIL 模型训练，图中纵坐标为训练过程中单步平均可靠度评分，横坐标为经历的迭代次数。

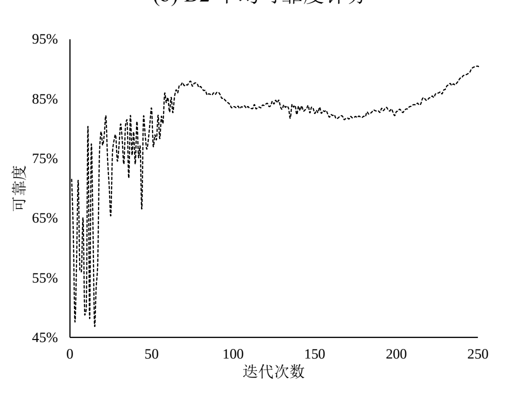
如图 6 所示，经过 250 次迭代后模型达到收敛，通过测试集测试模型，每次测试采样的样本数量为 200，采样 10 次得到的测试结果取平均，得到平均单步可靠度评分，结果如表 4 所示，三个数据集下模型的单步可靠度评分达到 89.30%、87.45%和 91.33%，相较于 IRL 和 BC 算法平均可靠度评分分别提升 9.16%和 15.84%，表明算法策略能够收敛于与被学习的人工策略具有较低的偏差的结果。



(a) D1 平均可靠度评分



(b) D2 平均可靠度评分



(c) D3 平均可靠度评分

图 6　不同数据集下模型训练平均可靠度评分

Fig. 6　Average model training reliability score in different datasets

为更好地掌握模型结果与人工数据之间的可靠度评分具体分布情况，采用 2000 条测试样本测试可靠度评分得到其分布，得到单步可靠度评分的分布如表 5 所示，可靠度评分在 90%以上的步数的占比分别达到 75.2%、72.6%和 84.4%，而可靠度评分低于 80%的区间包含步数的占比 13.85%、16.20%和 7.20%，说明人工分配方案在本模型结果中保持较高的可靠度。

表 4　平均单步可靠度评分统计

Tab. 4　Average reliability score of single step

	D_1	D_2	D_3	Average
FE-GAIL	89.30%	87.45%	91.33%%	**88.38%**
IRL	78.37%	76.64%	82.61%	79.22%
BC	73.35%	68.94%	74.83%	72.54%

表 5　单步可靠度评分分布

Tab. 5　Single step reliability score distribution

	D_1	D_2	D_3
91%-100%	75.20%	72.60%	84.35%
81%-90%	10.95%	11.20%	8.45%
<80%	13.85%	16.20%	7.20%

3.3　模型区分度结果分析

不同策略之间存在的差异将导致结果之间存在差异，为证明模型对于不同来源数据的区分度，本节通过 3 组不同数据集训练模型，分别代表不同的操作策略，测试模型对于各个数据集的鉴别效果。混淆矩阵是一种描述评价分类结果错误率的常用方式，其中根据模型结果正确与否和实际正反例之间的关系，将样本分为四个类别 TP、FN、FP、TN。样本 p 的所属标签类别 c_{sp} 可以通过设置不同的阈值 v_{th}，设通过数据集 D_i 训练得到的模型预测样本 p 标签值在采样数据中预测标签值排名为 v，即：

$$\begin{cases} c_{sp}=TP,\ \text{if}\ v>v_{th},\ p\in D_i \\ c_{sp}=NP,\ \text{if}\ v>v_{th},\ p\notin D_i \end{cases} \tag{14}$$

PR 曲线是通过计算不同的正例阈值下的精确率 P (Precision)和召回率 R (Recall)得到变化曲线，两者的计算公式如下：

$$P=\frac{TP}{TP+FP} \tag{15}$$

$$R=\frac{TP}{TP+FN} \tag{16}$$

为了综合考虑精确率和召回率，采用综合度量指标 F 度量值来评价其结果，其计算公式如下：

$$F=\frac{2PR}{P+R}=\frac{2TP}{2TP+FN+FP} \tag{17}$$

区分度测试集包含全部三组数据共 6000 个样本数据，分别采用三个模型对样本是否来自于该模型所对应的训练数据集进行判别，以此计算混淆矩阵。按照 10%为间隔分别选取不同的阈值 v_{th}，根据不同模型对应的混淆矩阵，得到不同训练数据集下 FE-GAIL 模型的 P-R 曲线如图 7 所示，图中被标注点表示各个模型取最大 F 值的点，说明在不同数据集下本模型的区分度性能接近，模型区分度性能受数据集影响小。

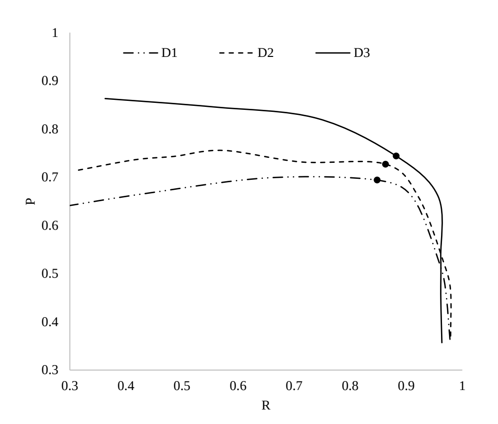


图 7　不同数据下模型 PR 曲线

Fig. 7　PR curves of the models under different data

图 7 中被标注的三个点为模型在三组数据下的最大 F 值，其值如表 6 所示，三个数据集的 F 值分别为 0.7632、0.7892 和 0.8071，说明模型对不同来源数据的区分能力较好。

表 6　不同数据下模型的 F 值

Tab. 6　F values of the models under different data

	D_1	D_2	D_3
FE-GAIL	0.7632	0.7892	0.8071
IRL	0.6743	0.6912	0.6843
BC	0.5958	0.6191	0.6367

3.4　算法同等参数误差水平分析

由于本算法是通过对神经网络进行参数学习的方式得到与人工决策相匹配的隐性参数设置，无法直接和人工参数配置进行相似度对比，因此将通过间接实验的方式测试算法同等参数误差水平。本实验中对三个数据集的生成过程加入一定的参数误差，测试参数误差下的分配结果与原始分配结果间的相似度，同时测试通过本算法进行参数学习所得的分配结果与原始分配结果的相似度，以获得本算法分配结果的同等参数误差水平。

定义相似度评分 q_{score} 的计算公式如下：

$$q_{score}=\alpha_1 N_s+\alpha_2 N_{sa}+\alpha_3 N_{dif} \quad (18)$$

其中与人工操作行为相比，N_s 为与其相同的动作数量，N_{sa} 为航班、机位属性 A_{fk} 和 A_{gi} 与其相同的动作数量，N_{dif} 为航班、机位属性 A_{fk} 和 A_{gi} 与其不同的动作数量，α_1、α_2、α_3 分别为三者对应的权重参数。

表 7 为三个数据集在不同参数误差下，生成的新数据集与原数据集以及通过原数据训练的模型得到的分配结果之间的对比情况。

表 7　不同参数误差的数据集

Tab. 7　Datasets of different parameter errors

数据集	参数集误差	N_s	N_{sa}	N_{dif}	q_{score}
D_1	5	10355	201	1444	0.7780
	10	9149	1088	1763	0.6185
	15	7007	2299	2694	0.1530
	20	6730	2479	2791	0.1045
	25	6418	2768	2814	0.0930
	FE-GAIL	7097	3709	1194	0.9030
	IRL	6488	3344	2168	0.416
	BC	5988	3246	2766	0.117
数据集	参数集误差	N_s	N_{sa}	N_{dif}	q_{score}
D_2	5	9328	460	2212	0.3940
	10	8227	1186	2587	0.2065
	15	7101	2260	2639	0.1805
	20	6749	2454	2797	0.1015
	25	6445	2748	2807	0.0965
	FE-GAIL	7663	2377	1960	0.5200
	IRL	6925	2546	2529	0.2355
	BC	6858	2287	2855	0.0725
数据集	参数集误差	N_s	N_{sa}	N_{dif}	q_{score}
D_3	5	10524	154	1322	0.8390
	10	8732	1418	1850	0.5750
	15	7058	2903	2039	0.4805
	20	6994	2931	2075	0.4625
	25	6638	3222	2140	0.4300
	FE-GAIL	7732	2981	1287	0.8565
	IRL	6985	2687	2328	0.336
	BC	6628	2442	2930	0.035

在三个数据集下，与 5%参数设置偏差下的结果相比，模型结果中 N_{dif} 的数量较少，但同时 N_s 的数量也较少，说明本模型的分配结果能够得到与人工数据属性特征更为接近的结果。对比 IRL 和 BC 算法模型，FE-GAIL 算法模型在不同数据集上的相似度评分均有较为明显的提升。FE-GAIL 模型的分配结果的相似度评分分别为 0.903、0.52、0.8565，分数较 5%参数设置偏差所带来的分配结果仍有优势，因此可以认为本算法进行参数学习后所得结果同等于 5%人工参数偏差下的分配结果。

4　结束语

针对机位再分配问题特性，提出了基于特征嵌入的生成对抗模仿学习算法从航班的机位再分配操作数据中学习其对应的操作策略的方法，用以解决航班延误时机场的停机位动态调整问题。在环境建模的过程中加入特征嵌入的表征方式提升机位和航班属性特性可读性，且由于 GAN 网络的加入缓解了动态环境中样本效率低下的问题，对比 IRL 和 BC 算法在停机位动态分配策略学习任务中多项指标均有较明显的优势。对多组不同的机场调度人员的调度策略数据进行了策略学习，对比不同参数误差情况下与原有操作数据之间的差异，证明通过本模型得到的结果可以规避 5%左右的参数设置偏差。本文研究方向为停机位动态分配，基于本文提出的 FE-GAIL 算法策略模型的机位分配模型输出指派方案具有与人工操作方式更高的一致性，可减少机位调度操作人员对算法推荐方案的二次调整，从提升算法结果的实用性的角度提升机场运行效率。

目前在被模仿策略的复杂度上，算法还有进一步的提升空间，后续的工作中可以加强算法对更为接近实际操作人员操作复杂性的数据进行策略拟合。

参考文献:

[1] Pternea M. Optimal Reassignment of Flights to Gates Focusing on Transfer Passengers [D] , 2019.

[2] Daş G S, Gzara F, Stützle T. A review on airport gate assignment problems: Single versus multi objective approaches [J]. Omega, 2020, 92: 102146.

[3] Deng Wu, Li Bo, Zhao Huimin. Study on an Airport Gate Reassignment Method and Its Application [J]. Multidisciplinary Digital Publishing Institute, 2017, 9 (11): 258.

[4] Zhang Dong, Klabjan D. Optimization for gate re-assignment [J]. Transportation Research Part B, 2017, 95: 260-284.

[5] Pternea M, Haghani A. Mathematical models for flight-to-gate reassignment with passenger flows: State-of-the-art comparative analysis, formu-lation improvement, and a new multidimensional assignment model [J]. Computers & Industrial Engineering, 2018, 123: 103-118.

[6] Pternea M, Haghani A. An aircraft-to-gate reassignment framework for dealing with schedule disruptions [J]. Journal of Air Transport Management, 2019, 78: 116-132.

[7] Yu Chuhang, Zhang Dong, Lau HYKH. A heuristic approach for solving an integrated gate reassignment and taxi scheduling problem [J]. Journal of Air Transport Management, 2017, 62: 189-196.

[8] 姜雨，胡志韬，童楚，刘振宇，陈丽丽，张洪海. 面向航班延误的停机位实时指派优化模型 [J]. 交通运输系统工程与信息, 2020, 20 (05): 185-190, 217. (Jiang Yu, Hu Zhitao, Tong Chu, *et al.* An Optimization Model for Gate Re-assignment under Flight Delays [J]. Journal of Transportation Systems Engineering & Information Technology, 20 (5): 185-190, 217.)

[9] Maharjan B, Matis TI, An optimization model for gate reassignment in

response to flight delays, Journal of Air Transport Management, 2011, 17 (4): 256-261.

[10] Chen Dian, Zhou Bragy, Koltun V, *et al.* Learning by cheating [C]// Conference on Robot Learning. PMLR, 2020: 66-75.

[11] Zhou Yang, Fu Rui, Wang Chang, *et al.* Modeling Car-Following Behaviors and Driving Styles with Generative Adversarial Imitation Learning [J]. Sensors, 2020, 20 (18): 5034.

[12] Lioutikov R, Neumann G, Maeda G, *et al.* Learning movement primitive libraries through probabilistic segmentation [J]. The International Journal of Robotics Research, 2017, 36 (8): 879-894.

[13] Finn C, Levine S, Abbeel P. Guided cost learning: Deep inverse optimal control via policy optimization [C]// International conference on machine learning. PMLR, 2016: 49-58.

[14] Sun W, Venkatraman A, Gordon G J, *et al.* Deeply aggrevated: Differentiable imitation learning for sequential prediction [C]// International Conference on Machine Learning. PMLR, 2017: 3309-3318.

[15] Pan Menghai, Huang Weixiao, Li Yanhua, *et al.* xGAIL: Explainable Generative Adversarial Imitation Learning for Explainable Human Decision Analysis [C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 1334-1343.

[16] Shou Zhenyu, Di Xuan, Ye Jieping, *et al.* Optimal passenger-seeking policies on E-hailing platforms using Markov decision process and imitation learning [J]. Transportation Research Part C: Emerging Technologies, 2020, 111: 91-113.

[17] Hussein A, Elyan E, Gaber M M, *et al.* Deep imitation learning for 3D navigation tasks [J]. Neural computing and applications, 2018, 29 (7): 389-404.

[18] Ho J, Ermon S. Generative adversarial imitation learning [J]. Advances in neural information processing systems, 2016, 29: 4565-4573.

[19] Torabi F, Warnell G, Stone P. Recent advances in imitation learning from observation [J]. arXiv preprint arXiv: 1905. 13566, 2019.

[20] Chen Xia, Kamel A E. Neural inverse reinforcement learning in autonomous navigation [J]. Robotics and Autonomous Systems, 2016, 84: 1-14.

[21] Vinyals O, Babuschkin I, Wojciech M. Czarnecki Grandmaster level in StarCraft II using multi-agent reinforcement learning [J]. Nature, 2019, Vol 575, 350-354.